# Warm Starting Series of Mixed-Integer Linear Programs

Sean Kelley

27 January 2022

# Overview

# Motivation

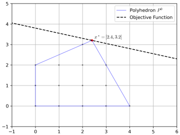
- Mixed-Integer Programming has created great value in industry.
- One important source comes from problems solved by a series of Mixed-Integer Linear Programs (MILPs).
  - Electric Grid Production Planning (Stochastic Dual Decomposition)
  - Vehicle Routing (Branch and Price)
- Many MILP instances in such series differ only by objective coefficients or bounds on their constraints.
- Solvers can leverage what they discover solving one MILP to more quickly solve a similar MILP (a.k.a. "Warm Start").
- Warm starting would enable greater performance and expand the space of tractable problems for those solved as a series of MILPs.

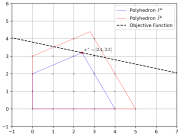
This presentation details potential opportunities to warm start series of MILPs and presents problem classes whose solutions are found by solving such series.

# Simple Example

$$\max\{x + 4y : -\frac{x}{2} + y \leq 2, \quad x + \frac{y}{2} \leq 4, (x,y) \in \mathbb{Z}^2_+\} \tag{1}$$

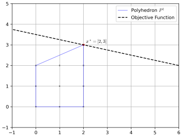
$$\max\{x + 4y : -\frac{x}{2} + y \leq 3, \quad x + \frac{y}{2} \leq 5, (x,y) \in \mathbb{Z}^2_+\} \tag{2}$$




Update bounds

Branch on $x \leq 2$ or $x \geq 3$ $\Longrightarrow$ Branch on $x \leq 2$ or $x \geq 3$



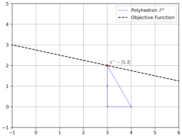
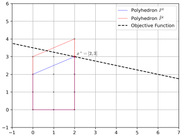




Figure 1: (1)'s Branch and Bound tree



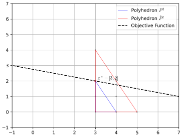




Figure 2: (2)'s Branch and Bound tree

After solving (1), we can get a tree for (2) without another solve by substituting (2)'s bounds in each subproblem of (1).

# Three Ways to Warm Start

After updating the objective and constraint bounds in all subproblems of Branch and Cut, we have the following three approaches to warm start.

(a) Good warm start:

- Recalculate the primal bound.
- Restart Branch and Cut from the root node.

(b) Better warm start:

- Parameterize cuts.
- Recalculate the primal and/or dual bounds.
- Restart Branch and Cut from the leaf nodes.

(c) Hypothesized best warm start:

- Parameterize cuts.
- Add to the root relaxation cuts valid for all terminal subproblems.
- Restart Branch and Cut from the root node.

# Valid Cuts for All Terminal Subproblems

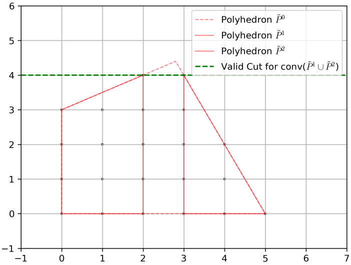



Figure 3: A cut valid for both of (2)'s subproblems

- We are motivated to explore (c) because:
  - It maintains (b)'s tightened relaxation while removing need to explore all subproblems.
  - It can remain effective when objective or constraint bounds change greatly.
- We want to find cuts such that:
  - No terminal subproblem is violated.
  - Some estimate of the solution is maximally violated.

We can find cuts valid for all terminal subproblems by formulating and solving a LP known as the Cut Generating LP (CGLP).

# Updating a Branch and Bound Tree for Warm Starting

- Let $t$ be the index of a node in a Branch and Bound tree.
- Define the feasible region of the LP subproblem in node $t$ as

$$P^t = \{x \in \mathbb{R}^n : A^t x \geq b^t, u^t \geq x \geq l^t\}$$

  where
  - $A^t = \begin{bmatrix} A \\ \Pi^t \end{bmatrix}$ and $b^t = \begin{bmatrix} b \\ \Pi_0^t(b) \end{bmatrix}$.
  - $b$ is the constraint bounds for the warm started MILP.
  - $(\Pi^t, \Pi_0^t(b))$ represent parameterized valid inequalities derived from cut generation routines of node $t$ and its ancestors in the previous solve.
  - $u^t$ and $l^t$ are bounds on $x$ from branching in the previous solve.
- Let $\mathcal{T}$ be the terminal nodes of a Branch and Bound tree.
- Cuts valid for all terminal subproblems are equivalent to those valid for

$$P_D = \text{conv}(\bigcup_{t \in \mathcal{T}} P^t)$$

# Deriving the Cut Generating LP

- Let $\mu^t \in \mathbb{R}^m_+$, $w^t \in \mathbb{R}^n_+$, $v^t \in \mathbb{R}^n_+$, and $I_n$ be the $n$-dimensional identity matrix.
- The following is a valid inequality for $P^t$:

$$A^{t\,T} \mu^t + I_n w^t - I_n v^t \geq b^{t\,T} \mu^t + l^{t\,T} w^t - u^{t\,T} v^t \tag{3}$$

- Let $\pi \in \mathbb{R}^n$ and $\pi_0 \in \mathbb{R}$ satisfy the following:

$$\left.\begin{aligned} \pi &\geq A^{t\,T} \mu^t + I_n w^t - I_n v^t \\ \pi_0 &\leq b^{t\,T} \mu^t + l^{t\,T} w^t - u^{t\,T} v^t \end{aligned}\right. \quad \text{for all } t \in \mathcal{T} \tag{4}$$

- Since (3) is a valid inequality for each LP in $\mathcal{T}$, it follows $\pi^T x \geq \pi_0$ is a valid inequality for all $x \in P_D$.

# Deriving the Cut Generating LP

We can generate a cut valid for all terminal subproblems that maximally violates a solution estimate by solving the CGLP.

- Let $\bar{x} \in \mathbb{R}^n$ be an estimate of the optimal solution the MILP instance.
- We define (CGLP) as follows:

$$
\begin{aligned}
\text{minimize} \quad & \pi^T \bar{x} - \pi_0 & & \\
\text{subject to} \quad & \pi \geq {A^t}^T \mu^t + I_n w^t - I_n v^t & & t \in \mathcal{T} \\
& \pi_0 \leq {b^t}^T \mu^t + {l^t}^T w^t - {u^t}^T v^t & & t \in \mathcal{T} \\
& 1 = \sum_{t \in \mathcal{T}} \Big( \sum_{j=1}^{m^t} \mu_j^t + \sum_{i=1}^{n} w_i^t + \sum_{i=1}^{n} v_i^t \Big) & & \\
& \mu^t \in \mathbb{R}_+^{m^t}, \; w^t \in \mathbb{R}_+^n \; v^t \in \mathbb{R}_+^n & & t \in \mathcal{T}
\end{aligned}
\qquad \text{(CGLP)}
$$

Warm start method (c) is completed by adding the solutions to (CGLP) as cuts to the root relaxation.

# Restarting MILP's

- Warm starting with a tightened root LP relaxation can be preferred to a partial tree already found on a looser root relaxation.
- We aim to identify problem classes where there is a strong preference for the former.
- When this is true, a partial solve followed by warm start method (c) takes less time than solving the instance cold.
- The idea then is as follows:
  - Solve LP relaxations for certain number of nodes.
  - Generate CGLP cuts against this tree.
  - Tighten root node with these cuts.
  - Restart Branch and Cut from beginning with tightened root relaxation.

If many problem classes exist where the above is beneficial, this would dramatically change how solvers work today.

# Dual Decomposition for Stochastic MILP

- Let $S^j := \{(x, y^j) : Ax \leq b, x \in X, T^j x + wy^j \leq h^j, y^j \in Y\}$
- A multistage stochastic MILP can be formulated as follows:

$$\max \left\{ c^T x + \sum_{j=1}^{r} p^j {q^j}^T y^j : (x, y^j) \in S^j \text{ for } j \in [r] \right\} \tag{5}$$

- It can be reformulated as follows:

$$\max \left\{ \sum_{j=1}^{r} p^j (c^T x^j + {q^j}^T y^j) : (x^j, y^j) \in S^j \text{ for } j \in [r], x^1 = \ldots = x^r \right\} \tag{6}$$

- For $H^j$ such that $\sum_{j=1}^{r} H^j x^j = 0$ is equivalent to $x_1 = \ldots = x_r$, (5) and (6) can have an upper bound decomposed as follows:

$$\min_{\lambda} D(\lambda) \tag{LD}$$

$$D(\lambda) = \max \left\{ \sum_{j=1}^{r} p^j (c^T x^j + {q^j}^T y^j) + \lambda (H^j x^j) : (x^j, y^j) \in S^j \text{ for } j \in [r] \right\}$$

# Dual Decomposition for Stochastic MILP

- $D(\lambda)$ consists of $r$ independent MILP's.
- (LD) is solved by a subgradient algorithm. We branch on the average value of $x$ in (LD)'s solution and bound (LD) again until $x^j$ are all identical.
- Notice there are $r$ series of MILPs sharing similar structure:
  - MILP $j$ in $D(\lambda)$ differs from MILP $j$ in $D(\lambda')$ only by objective.
  - MILP $j$ in $D(\lambda)$ before branching differs from MILP $j$ in $D(\lambda)$ after branching only by variable bounds.

Since solving a multistage Stochastic MILP involves solving multiple series of potentially difficult MILPs, the warm starting techniques presented earlier could be beneficial to the solution method.

# Warm Starting Big Picture

We have a couple approaches at our disposal for warm starting a MILP from a previously solved one.

(a) Good warm start:
- Recalculate the primal bound.
- Restart Branch and Cut from the root node.

(b) Better warm start:
- Parameterize cuts.
- Recalculate the primal and/or dual bounds.
- Restart Branch and Cut from the leaf nodes.

(c) Hypothesized best warm start:
- Parameterize cuts.
- Add to the root relaxation cuts valid for all terminal subproblems.
- Restart Branch and Cut from the root node.

# Next Steps

There are a few other problem classes we look to test the effectiveness of warm starts:

- Multiobjective MILP
- Primal Heuristics (RINS)
- Bilevel MILP
- Real-Time MILP
- MINLP

We have the following next tasks in mind:

- Make cut generation numerically safe.
- Parametrize cut generation.
- Implement algorithms for each problem class.
- Run experiments for each algorithm comparing warm started version to current algorithms.

We intend to test our warm starting methods against the problem classes mentioned in this presention to determine which can leverage these approaches to more effeciently be solved.

# Simple Approach to Warm Starting

- Warm starting means to reuse a Branch and Bound tree from solving a previous MILP when solving a new one.
- One approach is to reevaluate leaf nodes' solutions at the new MILP's objective and RHS to set applicable bounds. Then leaf nodes are placed into a priority queue, and Branch and Cut resumes.
- This approach has the following advantages:
  (a) When only the objective changes, all previously feasible solutions remain feasible. We can start the primal bound as the one with the best evaluation of the new objective.
  (b) When only the RHS changes, each node's dual LP relaxation maintains the same feasible region. We can start the dual bound of each node as its previous dual LP solution evaluated at the new RHS.
  (c) Restarting from the leaf nodes prevents us from having to branch and bound to regenerate them in the new instance.

# Simple Approach to Warm Starting

- Warm starting means to reuse a Branch and Bound tree from solving a previous MILP when solving a new one.
- One approach is to reevaluate leaf nodes' solutions at the new MILP's objective and RHS to set applicable bounds. Then leaf nodes are placed into a priority queue, and Branch and Cut resumes.
- This approach has the following disadvantages:
  (a) The primal or dual bound is weak if the objective or RHS changes significantly.
  (b) Potentially many leaf nodes will need to be branched.
  (c) Generated cuts may no longer be valid if RHS changed.

Starting a MILP solve with a Branch and Bound tree from a previous solve can be a simple and quick way to improve performance, but such an approach is not a panacea.

# Motivating the Cut Generating LP

- Generating strong cuts of the convex hull of the previous disjunctive terms' LP relaxations near the new optimal solution ameliorates disadvantages (a) and (b). Instead of reusing the previous tree to solve our problem, we solve with a new Branch and Bound tree with these cuts added to the root LP relaxation.
- This removes advantages (b) and (c) but tightens the root LP relaxation. In many cases this is preferred because it can more quickly give the solver an idea of where to search for an optimal solution.
- These cuts are the valid inequalities that violate by as much as possible our best estimates of the new optimal solution.
- Given such estimates, the Cut Generating LP (CGLP) finds such inequalities.

It may be possible to improve a warm start by replacing the previous tree with strategically chosen strong cuts of the convex hull of its disjunctive terms' LP relaxations.

# Handling Constraint Bound Changes

- Should $b^k \neq b^{k'}$, $(A^{kt}, b^{kt})$ may no longer be valid for the matching disjunctive term in MILP $k'$. Thus, MILP $k$'s tree may not be able to warm start MILP $k'$.
- Let $p_{k't}(A^{kt}, b^{kt}) = (p_{k't}(A, b^k), p_{k't}(\Pi^{kt}, \Pi_0^{kt})) : \mathbb{R}^{m^t \times n+1} \rightarrow \mathbb{R}^{m^t \times n+1}$ map the constraints $(A^{kt}, b^{kt})$ to constraints that are valid for matching disjunctive term in MILP $k'$, where
  - $p_{k't}(A, b^k) = (A, b^{k'})$
  - $p_{k't}(\Pi^{kt}, \Pi_0^{kt})$ updates each GMI cut as described in Guzelsoy (2006) and each CGLP cut $(\pi, \pi_0)$ as follows. Let the tree of MILP $\kappa$ be from which $(\pi, \pi_0)$ is derived. Let $(\tilde{A}, \tilde{b}) = p_{k't}(A^{\kappa t}, b^{\kappa t})$. Then

$$\begin{aligned}(\pi, \pi_0) =&(\max\{\tilde{A}^T \mu^{kt} + {I_n}^T w^{kt} - {I_n}^T v^{kt} \text{ for all } t \in \mathcal{T}^k\},\\ &\min\{\tilde{b}^T \mu^{kt} + {l^{kt}}^T w^{kt} - {u^{kt}}^T v^{kt} \text{ for all } t \in \mathcal{T}^k\})\end{aligned}$$

If $b^k \neq b^{k'}$, cuts generated for MILP $k$ must be made valid for MILP $k'$ before the tree of MILP $k$ can be used to warm start MILP $k'$.

# Branch and Price

- We can tighten the LP relaxation for a MILP with the following formulation:

$$\max_{x \in \text{conv}(\mathcal{S}_R)} \{c^T x | A'' x \leq b''\} \tag{7}$$

where

$$\mathcal{S}_R = \{x \in \mathbb{Z}^n | A' x \leq b'\}$$

- With $\mathcal{E}$ as the set of extreme points of conv($\mathcal{S}_R$), we can reformulate as follows:

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & \sum_{s \in \mathcal{E}} \lambda_s s = x \\ & A'' x \leq b'' \\ & \sum_{s \in \mathcal{E}} \lambda_s = 1 \\ & \lambda \in \mathbb{R}_+ \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \max \quad & \sum_{s \in \mathcal{E}} \lambda_s (c^T s) \\ \text{s.t.} \quad & \sum_{s \in \mathcal{E}} \lambda_s (A'' s) \leq b'' \\ & \sum_{s \in \mathcal{E}} \lambda_s = 1 \\ & \lambda \in \mathbb{R}_+ \end{aligned} \tag{DWLP}$$

# Branch and Price

- We solve (DWLP) by column generation, only ever using a subet of $\mathcal{E}$. For dual values $u$ and $\alpha$ of (DWLP)'s constraints, we find a new $s$ to augment the current subset by solving the following:

$$-\alpha + \max_{x \in \mathcal{S}_R} \{(c - uA'')x\} \qquad \text{(LR}(u))$$

- (DWLP) is often the LP relaxation in Branch and Price. Notice (LR($u$)) represents a series of closely related MILP's.
  - Within a single solve of (DWLP), each (LR($u$)) differs only by objective.
  - A common branching technique in Branch and Price is to bound variables in (LR($u$)). Thus, for fixed $u$, (LR($u$)) differs from its ancestor nodes by its RHS.

By warm starting the pricing problem in Branch and Price, we may be able to expand the sets of problems solved by this algorithm.

# Warm Starting Big Picture

- There are a couple of different approaches we can take to warm starting a MILP $k'$ from one previously solved, MILP $k$.
- All look to accomplish the same thing, but they differ from each other in the following ways:
  - When solving MILP $k'$ with the tree from MILP $k$'s solve and only the objectives differ between the two instances, there is no need to transform cuts, and only the feasible leaves need searched. However, the dual function is lost, so search priority is less clear.
  - When solving MILP $k'$ with the tree from MILP $k$'s solve and only the RHS's differ between the two instances, cuts may need to be transformed to remain valid, and all leaves need searched. However, the dual function remains intact, so search priority is clearer.
  - When generating CGLP cuts from MILP $k$ to warm start MILP $k'$, a partially searched tree is exchanged for a tighter root relaxation. Some nodes may be re-evaluated, but the number of nodes to search may be reduced.